

ABSTRACT

A method and apparatus are described for data synchronization between a client and a repository. According to one embodiment, data synchronization between a client and a repository is performed based on the results of a comparison between message digests associated with files stored on the client and a database of message digests stored on the repository. The message digests generated on the client uniquely identify the content of files stored on the client. This unique identification of the contents of the files on the client is accomplished by performing a cryptographic hash of the contents of the individual files. The database of message digests stored on the repository contains message digests from clients that are stored in the database at the time of data synchronization. By comparing message digests generated on the client with those stored on the repository, the need for data synchronization may be efficiently determined.